Joint transfer component analysis and metric learning for person re-identification

Yixiu Liu, Yunzhou Zhang✉, Sonya Coleman and Jianning Chi

A novel and efficient metric learning strategy for person re-identification is proposed. Person re-identification is formulated as a multi-domain learning problem. The assumption that the feature distributions from different camera views are the same is overthrown in this Letter. ID-based transfer component analysis (IDB-TCA) is proposed to learn a shared subspace, in which the differences in the feature distribution between source domain and target domain are significantly reduced. Experimental evaluation on the CUHK01 dataset demonstrates that metric learning with IDB-TCA embedded outperforms state-of-art metric methods for person re-identification.

Introduction: Person re-identification, aiming to finding the images that match the target person in a large-scale image library, greatly reduces the time cost of human search. Due to its great significance to visual supervision, it has rapidly become a research hotspot in the field of computer vision in recent years. A person usually undergoes large variations of appearance, scale, occlusion, posture and view angle under several different camera views, making person re-identification a valuable yet challenging research topic. Previous works on these challenges mainly concentrate on feature representation or metric learning. We are committed to the second methods in this Letter.

Recently, many effective metric methods have been proposed successively. The famous KISSME [1] method was proposed to measure the similarity of two samples using Mahalanobis distance. On the basis of KISSME, cross-view quadratic discriminant analysis (XQDA) was proposed in [2]. It extended KISSME algorithm to cross-view metric learning to further improve the matching rate of person re-identification. XQDA was based on the assumption that the feature distributions of two different camera views (source domain and target domain) were consistent with the Gaussian distribution.

In this Letter, the above assumption is no longer valid. The significant difference between our method and conventional metric learning methods is that we recognise that the feature distributions of different camera views are different. First, ID-based transfer component analysis (IDB-TCA) is proposed to find a transformation from current feature space to Hilbert space, in which the feature distribution is almost the same. Different from the original TCA [3], we maintain the principle that the feature distribution of pedestrians who have the same ID remain consistent across different camera views. Second, metric learning method is executed to maximise the sample distance between the classes and minimise the sample distance within the class. The proposed method is illustrated in Fig. 1 and detailed below.

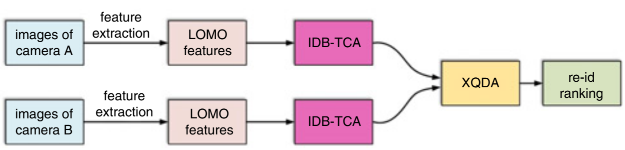


Fig. 1 *System overview. Modules of IDB-TCA are our novel contributions*

Assumption analysis: The labelled data pairs in the source and target domain are, respectively, defined as $D_S = \{(x_1, y_1), \ldots, (x_{n1}, y_{n1})\}$ and $D_T = \{(x_1, y_1), \ldots, (x_{n2}, y_{n2})\}$, where $\boldsymbol{x}_i$ denotes the feature vector and y_i is the corresponding ID label. $\mathcal{P}(X_S)$ and $\mathcal{Q}(X_S)$ (or $\mathcal{P}$ and $\mathcal{Q}$ in short) denote the marginal distributions of $X_S = \{x_1 \ldots x_{n1}\}$ and $X_T = \{x_1 \ldots x_{n2}\}$ from the source and target domains, respectively.

Conventional metric learning methods for person re-identification consider that $\mathcal{P} = \mathcal{Q}$, which is not completely objective. For example, the feature distributions of three samples from different camera views in a public person re-identification dataset CUHK01 are shown in Fig. 2*a*. From the curve, we can clearly see that there are obvious differences in the feature distribution of the same person between the source and target domain.

In this Letter, we consider that $P \neq Q$, our work is to learn a transformation ϕ to transfer the features from current space to Hilbert space such that $P(\phi(X_S)) \approx P(\phi(X_T))$. After the transformation, the corresponding feature distributions of three samples of Fig. 2*a* are shown in Fig. 2*b*. The two curves that represent the same IDs are almost overlapping, indicating that the distribution differences are largely eliminated.

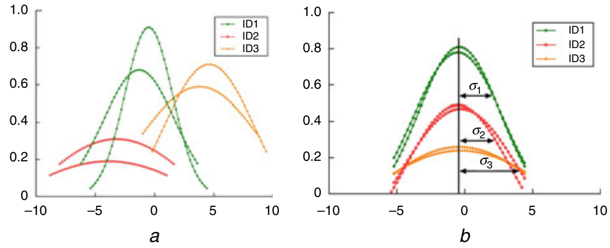


Fig. 2 *Effect of IDB-TCA on reducing the differences of feature distribution*

IDB-TCA: Based on the original TCA algorithm [3], IDB-TCA is proposed in this Letter. The new method is more restrictive. It requires that the feature distributions with the same ID in the source and target domains are almost the same.

IDB-TCA learns a transformation by solving an optimisation problem, so that the feature distributions in the source domain and target domain are almost the same in the transformed subspace. The object function is formulated according to maximum mean discrepancy as

$$\text{dist}^2(D_S, D_T) = \sum_{l=1}^{N} \left\| \frac{1}{n_{Sl}} \sum_{\boldsymbol{x}_i \in D_{Sl}} \boldsymbol{A}_l^T \boldsymbol{x}_i - \frac{1}{n_{Tl}} \sum_{\boldsymbol{x}_j \in D_{Tl}} \boldsymbol{A}_l^T \boldsymbol{x}_j \right\|_H^2, \tag{1}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the feature vectors, $\|\cdot\|_H^2$ denotes the reproducing kernel Hilbert space norm, $\boldsymbol{A}$ is the projection matrix and $l \in \{1, \ldots, N\}$ means the ID number of pedestrians in dataset. Combined with the PCA algorithm, (1) can be expressed as

$$\text{dist}^2(D_S, D_T) = \sum_{l=1}^{N} \text{tr}(\boldsymbol{A}_l^T \boldsymbol{X}_l L_l \boldsymbol{X}_l^T \boldsymbol{A}_l), \tag{2}$$

where $\boldsymbol{X} = \{X_{Sl}, X_{Tl}\}$ is a feature matrix, $H = I - (1/n)qq^T$, $\boldsymbol{q} \in R^n$ is a column vector whose element is 1, $n = n_S + n_T$, and the element of $L \in R^{n \times n}$ can be expressed as

$$L_{i,j} = \begin{cases} \dfrac{1}{n_{Sl} n_{Sl}} & (\boldsymbol{x}_i, \boldsymbol{x}_j \in D_{Sl}, i, j \in [1, 2, \ldots, n]), \\ \dfrac{1}{n_{Tl} n_{Tl}} & (\boldsymbol{x}_i, \boldsymbol{x}_j \in D_{Tl}, i, j \in [1, 2, \ldots, n]), \\ \dfrac{-1}{n_{Sl} n_{Tl}} & (\text{others}) \end{cases} \tag{3}$$

Finally, the optimisation problem of IDB-TCA is formulated as

$$\begin{aligned} &\min_{\boldsymbol{A}_l} \text{dist}^2(D_S, D_T) = \min_{\boldsymbol{A}_l} \sum_{l=1}^{N} \text{tr}(\boldsymbol{A}_l^T \boldsymbol{X}_l L_l \boldsymbol{X}_l^T \boldsymbol{A}_l) \\ &\text{s.t. } \boldsymbol{A}_l^T \boldsymbol{X}_l H_l \boldsymbol{X}_l^T \boldsymbol{A}_l = I \end{aligned} \tag{4}$$

Metric learning in subspace: After the transformation, XQDA performs in the new space. We define the mapping of $\boldsymbol{X}$ in this space as Z (specifically, $\phi(X) = Z$). The distance function between z_i and z_j can be expressed as

$$d(z_i, z_j) = (z_i - z_j)^T \left(\sum\nolimits_I^{-1} - \sum\nolimits_E^{-1}\right)(z_i - z_j), \tag{5}$$

where $\sum_I$ and $\sum_E$ denote the covariance matrices of intrapersonal variations Ω_I and extrapersonal variations Ω_E, respectively.

XQDA uses the cross-view data to train and obtain a r dimensional subspace $\Psi \in R^{d \times r}$, in which a distance function is learned to measure the similarity of the two samples. Given the subspace Ψ, this distance function (5) can be expressed as

$$d_\Psi(z_i, z_j) = (z_i - z_j)^T \Psi \left(\sum\nolimits_I^{-1} - \sum\nolimits_E^{-1}\right) \Psi^T (z_i - z_j) \tag{6}$$

Through the IDB-TCA algorithm, Ω_I and Ω_E centre at zero. Given a basis φ, the projected samples of the two classes will still have the same mean but different variances. In this case, the variances $\sigma_I(\varphi)$

and $\sigma_E(\varphi)$ can be used to distinguish the two classes. Notice that $\sigma_I(\varphi) = \varphi^T \Sigma_I \varphi$ and $\sigma_E(\varphi) = \varphi^T \Sigma_E \varphi$, therefore, φ can be optimised such that $\sigma_E(\varphi)/\sigma_I(\varphi)$ is maximised as shown in the following equation:

$$G(\varphi) = \frac{\varphi^T \sum_E \varphi}{\varphi^T \sum_I \varphi} \tag{7}$$

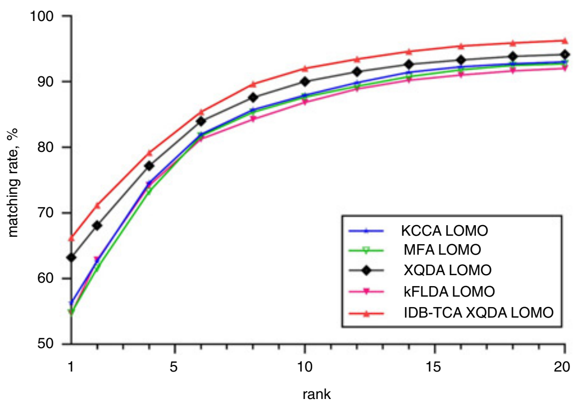
The maximisation of $G(\varphi)$ is equivalent to

$$\min_{\varphi} \ \varphi^T \sum\nolimits_E \varphi, \quad \text{s.t.} \ \ \varphi^T \sum\nolimits_E \varphi = 1, \tag{8}$$

with $\Psi = (\varphi_1, \varphi_2, \ldots, \varphi_r)$, we can learn a discriminant subspace and a distance function, as defined in (6).

Experimental results: The new algorithm is validated on the CUHK01 [4] dataset. It contains 3884 pedestrian images, as well as 971 pedestrian IDs. There are four images taken by one pair of disjoint cameras for each identity. Each identity has two images under the same camera area. All images are scaled to 160×60 pixels. The cumulative matching characteristic (CMC) curve is utilised to validate the method by finding the correct match in the top k matches. The feature extraction settings are the same as [2]. 35722D-features extracted from each image are reduced to 3400D by the IDB-TCA algorithm. In addition, the parameter is set as $N = 971$.

In the first experiment, we compare the proposed method with other methods which are directly performing metric learning in the local maximal occurrence (LOMO) [2] feature space, including kernel canonical correlation analysis (KCCA) [5], XQDA and kernel local Fisher discriminant analysis [6]. In order to make fair comparisons with them, the widely adopted experimental protocol on the CUHK01 dataset is to randomly divide the images into two equal parts, one for training and the other for testing. The procedure is repeated ten times with different random dataset splits to get an average performance. The CMC results are shown in Fig. 3*a*. As is shown, the proposed algorithm (LOMO + IDB-TCA XQDA) is superior to other algorithms since it reduces the differences of feature distribution between the source and target domain.



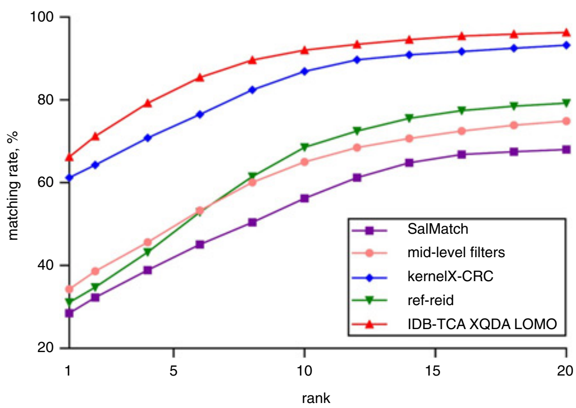


Fig. 3 *CMC curves for CUHK01 dataset*
a LOMO-related methods
b Current methods

In the second experiment, we compare the proposed method with the existing state-of-the-art methods, including mid-level filters [7], kernel cross-view collaborative representation based classification (kernel X-CRC) [8] and person re-identification with reference descriptor (Ref-reid) [9]. The CMC results are shown in Fig. 3*b*. We can see that the proposed method achieves the best result, as compared with other existing methods. Specifically, compared with kernel X-CRC, the new method increases from ∼61.20 to 66.23% at rank-1 and from 73.26 to 80.63% at rank-5.

Conclusion: A novel IDB-TCA method for metric-based person re-identification has been proposed. On the basis of original TCA algorithm, a new limitation is added that the differences of the feature distributions between different camera views from the same IDs are minimised. Thanks to that, the performance of the following metric learning has been obviously improved, which further improve the matching rate of person re-identification. The proposed method has been validated on the CUHK01 dataset and the results show that the new method outperforms many state-of-the-art methods.

Acknowledgments: The research is supported by the National Natural Science Foundation of China (61471110, 61733003), National Key R&D Program of China (2017YFC0805000/5005), Fundamental Research Funds for the Central Universities (N172608005, N160413002).

Submitted: *5 February 2018* E-first: *29 May 2018*
doi: 10.1049/el.2018.0324
One or more of the Figures in this Letter are available in colour online.

Yixiu Liu (*College of Information Science and Engineering, Northeastern University, Shenyang 110819, People's Republic of China*)

Yunzhou Zhang and Jianning Chi (*Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, People's Republic of China*)

✉ E-mail: zhangyunzhou@mail.neu.edu.cn

Sonya Coleman (*Intelligent Systems Research Centre, University of Ulster, Londonderry, United Kingdom*)

References

1 Köstinger, M., Hirzer, M., Wohlhart, P., *et al.*: 'Large scale metric learning from equivalence constraints'. 2012 IEEE Conf. on Computer Vision and Pattern Recognition, Providence, RI, USA, June 2012, pp. 2288–2295
2 Liao, S., Hu, Y., Zhu, X., *et al.*: 'Person re-identification by local maximal occurrence representation and metric learning'. 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015, pp. 2197–2206
3 Pan, S.J., Tsang, I.W., Kwok, J.T., *et al.*: 'Domain adaptation via transfer component analysis', *Trans. Neural Netw.*, 2011, **22**, (2), pp. 199–210
4 Li, W., Zhao, R., and Wang, X.: 'Human reidentification with transferred metric learning'. Asian Conf. on Computer Vision, Daejeon, Korea, November 2012, pp. 31–44
5 Lisanti, G., Masi, I., and Bimbo, A.D.: 'Matching people across camera views using kernel canonical correlation analysis', 2014, pp. 1–6
6 Xiong, F., Gou, M., Camps, O., *et al.*: 'Person re-identification using kernel-based metric learning methods'. European Conf. on Computer Vision, Zurich, Switzerland, September 2014, pp. 1–16
7 Zhao, R., Ouyang, W., and Wang, X.: 'Learning mid-level filters for person re-identification'. Computer Vision and Pattern Recognition, Columbus, OH, USA, June 2014, pp. 144–151
8 Prates, R., and Schwartz, W.R.: 'Kernel cross-view collaborative representation based classification for person re-identification', 2016, https://arxiv.org/abs/1611.06969
9 An, L., Kafai, M., Yang, S., *et al.*: 'Person reidentification with reference descriptor', *Trans. Circuits Syst. Video Technol.*, 2016, **26**, (4), pp. 776–787